International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Tasks Scheduling with Virtual Machines of the Deadline-Aware Priority Scheduling Model in Cloud Computing

Arvind Kumar Singh[1], Dr. Hitendra Singh[2], Manish Varsney[3]

Submitted: 09/10/2023 **Revised**: 28/11/2023 **Accepted**: 09/12/2023

Abstract*:* Cloud computing involves providing applications as services over the Internet, encompassing both the software applications provided as Software as a Service (SaaS) and the hardware and systems software housed in datacenters supporting these services. An Open Cloud, offering utility computing services, is characterized by its availability to the public on a pay-as-you-go model. The Cloud Sim toolkit serves as a tool for configuring and optimizing policies across all Cloud Sim components, making it a valuable resource for studying the complexities arising from diverse scenarios. For simulation tests, a laptop with the following specifications will be employed: a 2.5 GHz Intel Core i5 processor, 4 GB of RAM, and a 512 GB hard drive. The DAPS model adopts a methodology where tasks are prioritized in ascending order based on length, and the state of the Virtual Machine (VM) is deemed successful if it meets the deadline constraint. Subsequently, various jobs are allocated to the appropriate VM's, aiming to minimize the makespan and completion time. Experimental results indicate that the proposed approach outperforms existing techniques by reducing average makespan while enhancing the diversity of tasks

Keywords: *Task scheduling, Cloud Computing, Load, balancing, Priority Scheduling model****.***

1. Introduction

Cloud Computing(CC) includes both web-delivered applications and the underlying hardware and systems software found in data centers that facilitate these applications. The services, traditionally known as Software as a Service (SaaS), are augmented by the data center's hardware and software, collectively referred to as a Cloud. If this Cloud is provided to the general public on a pay-as-you-go model, it is termed an Open Cloud, and the corresponding service is referred to as Utility Computing. On the other hand, the term Private Cloud is used to characterize the internal data centers of a business or organization that are not accessible to the public. Therefore, Cloud Computing includes SaaS and Utility Computing while excluding Private Clouds. Individuals can take on roles as either users or providers of SaaS, as well as users or providers of Utility Computing. Cloud computing represents a novel technology employed to deliver various services [1][2].

Web-delivered apps and the hardware and systems software found in datacenters that enable them are both included in the concept of cloud computing. In addition to the services themselves, which have long been known as Software as a Service (SaaS), there is also datacenter hardware and software, which are referred to as a "Cloud." This Cloud is called an Open Cloud and the related service is called Utility Computing when it is made available to the public on a pay-per-use basis. If a company or organization's internal datacenters are kept private from the general public, they are referred to as private clouds. SaaS and utility computing are thus included in cloud computing, while private clouds are not. People can operate as SaaS consumers or suppliers or as Utility Computing users or providers. One cutting-edge technology used to provide a range of services is cloud computing [3][4].

1.1 Load balance awareness in task scheduling

Another significant restriction in cloud settings is load balancing, which has a significant impact on the performance of the task scheduling process. Even during task scheduling process, it is necessary to ensure that the resources are neither overloaded nor under loaded in any way. Moreover, as the majority of resources are projected to be distributed concurrently, the scheduling system must make sure that the distribution process maintains a balanced distribution of resources in relation to its capacity. Due to the fact that the issue is not only overcrowded VMs, but also under loaded idle VMs, this helps to improve resource usage. When virtual machines (VMs) remain in an idle state, it has an impact on the service provider's earnings. Hence, to optimize the utilization of existing resources, it is essential to incorporate an efficient load-balancing-aware job scheduling algorithm. While various methods can be employed for load balancing, dynamic load balancing is often favored over static load balancing, especially given

[1]*Research Scholar, Dept. of Computer Science,*
Maharishi University of Information Technology, Lucknow
E-mail id-arvindsinghfet@gmail.com
[2]*Assistant Professor, School of Engineering and Technology,*
Maharishi University of Information Technology, Lucknow
[3]*Professor, School of Engineering & Technology,*
Maharishi University of Information Technology, Lucknow

the real-time orientation of cloud computing architecture.When it comes to dynamic load balancing, there are two options: one is that it may be performed during the task scheduling process, and another is that it can be performed after the task scheduling process [5][6].

2. Literature Review

Gagandeep Kaur (2021) Cloud computing has gotten famous in these days as it gives on-request accessibility of PC resources like processor, stockpiling, and data transfer capacity all through the world. Virtualization helps in the development of data and calculation focuses and makes responsibility adjusting a lot more straightforward and simpler. Virtualization in current days makes the things simpler for cloud clients as they at this point don't have to make enormous forthright speculations and get quick admittance to minimal expense and adaptable resources. Virtualization intends to decrease end-client worries concerning worker upkeep, limitations, and adaptability. Cloud computing utilizes the idea of virtualization and utilizations IT resources as utilities in todays' reality. Cloud computing utilizes cloud commercial center to get to and incorporate the administrations and contributions. Cloud clients and cloud suppliers have various points while dealing with the resources, strategies, and request designs in true situation. In this paper, three-level engineering has been introduced to deal with the resources in financial cloud market considering cutoff time and execution season of the errands for clients.

Shaw, Rachael (2021) Cloud-based services have been widely adopted by a variety of businesses in the last several years, with a significant growth in usage. The goal is to cut expenses while improving service dependability and flexibility. But as cloud computing becomes more and more common, worries about data centers' excessive energy use and related carbon emissions have grown. Inefficient resource management is a significant contributor to this high energy consumption. While considerable efforts have been directed toward developing more advanced and energy-efficient resource management strategies, the presence of execution uncertainty poses a substantial obstacle for cloud resource management systems.

Harvinder Singh et al (2020) Cloud resource management is pivotal for proficient resource allotment and booking that needs for satisfying clients' assumptions. However, it is hard to foresee a fitting coordinating in a heterogeneous and dynamic cloud climate that prompts execution corruption and SLA infringement. Hence, resource management is a difficult assignment that might be compromised in light of the unseemly portion of the necessary resource. This paper presents a precise audit and insightful correlations of existing studies, research work exists on SLA, resource designation and resource planning for cloud computing. Further, conversation on open exploration issues, ebb and flow status and future examination headings in the field of cloud resource management.

Mohd Ameen Imran et al (2020) Cloud computing is a platform that is becoming used these days. Because of the use of designs such as SASS, PASS, and IASS in cloud computing, we have gained numerous advantages over our current platforms, including economies of scale and availability, security as well as other significant enhancements to our computing platforms. Many different types of research are being conducted to improve the reliability of cloud computing platforms for users (individual or corporate) and customers. This research work examines log managing in CC and demonstrates how the logs may be utilized as a useful data source on CC platforms such as Microsoft Azure, AWS, GCP, and others. It is proposed that cloud platforms preserve log files in permanent storage memory in a consistent format, which may be used for VM restoration and account monitoring for faults, as well as for the forensics process. This framework is called Lass scheme. Cloud platforms employ different types of services, and Lass provides a framework for collecting logs from multiple sources based on the type of service used. Loss provides a mechanism through which the user's log may be secured and the confidentiality of the user's log record can be maintained [7][8][9].

3. Objectives

- To awareness of load balance in task scheduling.
- To experiments with diverse Virtual Machines (VMs) within the DAPS model.

4. Research Methodology

4.1 Simulation Tool

An efficient tool for simulating task scheduling in homogeneous and heterogeneous cloud computing environments will be required. This tool should be flexible enough to support this environment while also receiving an increasing number of user requests. The toolkit will be a simulator that operates on an event-driven model and will be hosted in the Java programming language. The Cloud Sim toolkit will enable you to design and enhance the policies that will be applied across all of the Cloud Sim components. As a result, it will be regarded as a useful study tool since it can replicate the complexity that emerges from different settings. In order to conduct the simulation tests, the following setups will be used on a laptop: 2.5 GHz Intel Corei5 processor with 4 GB of RAM, and a 512 GB hard drive [10][11].

4.2 Datasets

The datasets that will be used in the cloud computing environment vary depending on the applications that are being utilized. It will be used in this study to build two different kinds of datasets: one that will be generated by using the Cloud Sim toolkit, and the other that will be collected from the NASA research webpage [12][13].

4.3 Proposed Models

The Deadline-Aware Priority Scheduling (DAPS) model will be the first scheduling model used. DAPS aims to plan and allocate tasks to suitable resources, thereby reducing the time needed for their completion within the deadline constraint and minimizing the minimum completion time. When crafting an efficient task scheduling model, it becomes crucial to consider the users' satisfaction level with the system as one of the variables. The approach involves focusing on scheduling activities based on a deadline constraint, where the deadline is regarded as the time necessary to accomplish the work [14][15].

5. Result and Discussion

Tasks were produced and separated into groups of 250, 500, 1000, 1500, and 2000 tasks, with each block consisting of 500 tasks. We used the subsequent metrics to evaluate the proposed DAPS model: the average of make span, the average of total average response time, resource utilization, guarantee ratio, and the number of breaches, as shown in Table 1. [16][17].

Table 1: Experiments with multiple DAPS model VMs for jobs

Experiments	Count of VMs	Count of tasks	Avg. make span	Mean of (total average RT)	Resource utilization	Guarantee ratio	No. of violations	Violation ratio
1	6	250	262818	34490	5621	0.992	2	0.8%
2	8	500	380478	45741	8403	0.99	5	1%
3	10	1000	568431	73229	14355	0.989	11	1.1%
4	12	1500	696511	88045	18389	0.992	12	0.8%
5	14	2000	716719	85337	19747	0.992	16	0.8%

5.1 Performance Assessment

In comparison to existing scheduling algorithms, the suggested DAPS model provides a better task scheduling approach based on a deadline constraint. The DAPS model utilized multiple performance metrics to minimize both the average makespan and the number of violated tasks simultaneously. Within this chapter, we examined the following five experiments to evaluate the overall quality and dependability of the proposed DAPS model [18].

i. 250 tasks with six VMs.

ii. 500 tasks with eight VMs.

iii. 1000 tasks with ten VMs.

iv. 1500 tasks with twelve VMs.

v. 2000 tasks with fourteen VMs

Using measures such as average of makespan, average of total average response time, average of total average response time, resource utilisation, guarantee ratio, amount of violations and violation ratio, we evaluate our results DAPS model to the GA, Min-Min, SJF, and Round Robin algorithms.

A crucial criterion and primary objective of the DAPS model we propose is the reduction of makespan, signifying the total time required for the completion of all tasks within a virtual machine. Table 2 demonstrates that the suggested DAPS model outperformed other algorithms in determining the overall makespan, particularly with an increased number of tasks when compared to GA, Min-Min, SJF, and Round Robin (RR) algorithms. The outcomes of all experiments consistently indicated the superior performance of the proposed model in reducing makespan while diversifying the range of tasks.

Table 2: The average makespan of the RR, SJF, GA, Min-Min, and DAPS

Experiments		Average makespan				
Count of tasks	Count of VMs	RR	SJF	GA	Min-Min	DAPS
250	6	323316	307421	305004	280130	262818
500	8	474695	440850	430264	422428	380478
1000	10	630876	645316	626303	598599	568431
1500	12	751648	785707	763337	733366	696511
2000	14	826665	814443	795613	760132	716719

In Figure 1, the X-axis denotes the quantity of tasks, while the Y-axis illustrates the increase in average makespan over time. When compared to existing algorithms, the suggested DAPS model has achieved a shorter average makespan in all of the trials conducted.

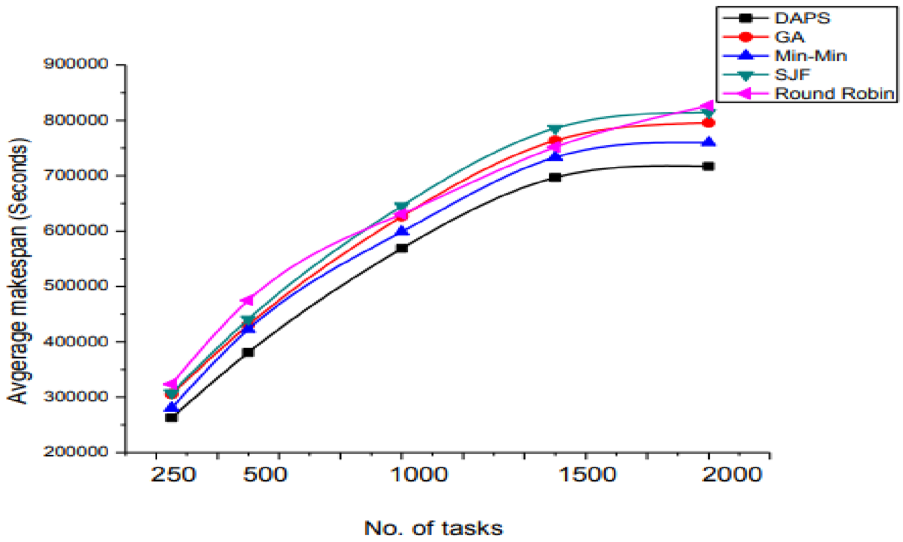

Fig 1: Contrasting of average makespan

The DAPS model we propose demonstrates a significant enhancement in makespan ratio, surpassing GA, Min-Min, SJF, and Round Robin by 10%, 6%, 12%, and 12.7%, respectively. Moreover, a statistical study employed the T-test to compare the DAPS model with other algorithms, including GA, Min-Min, SJF, and RR

6. Conclusion

The DAPS model employs a method where tasks are prioritized in ascending order based on length, and the VM's state is estimated successful upon meeting the deadline restriction. Subsequently, jobs are allocated to the appropriate VM while minimizing both the makespan and completion time. The overarching goal of the DAPS model is to achieve optimal execution by reducing system of measurement such as average makespan, mean of the overall average response time, number of violations, and violation percentage, ultimately ensuring user satisfaction. Furthermore, the The DAPS approach aims to maximize guarantee ratio and resource usage, particularly when the user is primarily concerned with the cost of task completion. This cost encompasses processing, memory, bandwidth, and storage, all within the constraints of the budget. The model's approach is designed to strike a balance between performance optimization and resource cost-effectiveness to meet user expectations.

References

[1] Gagandeep Kaur (2021). Framework for Resource Management in Cloud Computing. 10.1007/978-981-15-7062-9_3.

[2] Shaw, Rachael (2021),” Applying machine learning towards automating resource management in cloud computing environments”,

[3] Harvinder Singh et al (2020),” Cloud Resource Management: Comparative Analysis and Research

Issues”, International Journal of Scientific & Technology Research Volume 9, Issue 06

[4] Mohd Ameen Imran et al (2020),” Log as a Secure Service Scheme (LASS) for Cloud”, Journal of Scientific Research

[5] Singh, Saurabh, Young-Sik Jeong, and Jong Hyuk Park. "A survey on cloud computing security: Issues, threats, and solutions." Journal of Network and Computer Applications 75 (2016): 200-222.

[6] Gao, Yue, et al. "An energy and deadline aware resource provisioning, scheduling and optimization framework for cloud systems." Hardware/Software Codesign and System Synthesis (CODES+ ISSS), 2013 International Conference on. IEEE, 2013:1-10.

[7] Dong, Ziqian, Ning Liu, and Roberto Rojas-Cessa. "Greedy scheduling of tasks with time constraints for energy-efficient cloud-computing data centers." Journal of Cloud Computing 4.1 (2015): 1-14.

[8] Xu, Minxian, Wenhong Tian, and Rajkumar Buyya. "A survey on load balancing algorithms for virtual machines placement in cloud computing." Concurrency and Computation: Practice and Experience 29.12 (2017):1-20.

[9] Deldari, Arash, Mahmoud Naghibzadeh, and Saeid Abrishami. "CCA: a deadline-constrained workflow scheduling algorithm for multicore resources on the cloud." The journal of Supercomputing 73.2 (2017): 756-781.

[10] Singh, Sukhpal, Inderveer Chana, and Maninder Singh. "The Journey of QoSAware Autonomic Cloud Computing." IT Professional 19.2 (2017): 42-49.

[11] Panda, Sanjaya K., Indrajeet Gupta, and Prasanta K. Jana. "Task scheduling algorithms for multi-cloud systems: allocation-aware approach." Information Systems Frontiers (2017): 1-19.

[12] Narayan, Vipul, et al. "A Comprehensive Review of Various Approach for Medical Image Segmentation and Disease Prediction." Wireless Personal Communications 132.3 (2023): 1819-1848.

[13] Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classification: AN APPROACH TOWARD TAGGING UNLABELED MEDICAL DATASET." Journal of Scientific & Industrial Research (JSIR) 82.08 (2023): 818-830.

[14] Narayan, Vipul, et al. "Severity of Lumpy Disease detection based on Deep Learning Technique." 2023 International Conference on Disruptive Technologies (ICDT). IEEE, 2023.

[15] Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215.

[16] Kumar, Vaibhav, et al. "A Machine Learning Approach For Predicting Onset And Progression"“Towards Early Detection Of Chronic Diseases “." Journal of Pharmaceutical Negative Results (2022): 6195-6202.

[17] Chaturvedi, Pooja, Ajai Kumar Daniel, and Vipul Narayan. "Coverage Prediction for Target Coverage in WSN Using Machine Learning Approaches." (2021).

[18] Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.